

AMENDMENTS TO THE CLAIMS

This listing of claims will replace all prior versions and listings of claims in the application:

1. (Currently amended) A mathematical expression recognizing device comprising:
 - a character recognition unit configured to recognize characters in a document image containing a text and a mathematical expression;
 - a first dictionary configured to store a pair of evaluation scores for each type of word that can be identified by means of normal expression, a first one of the pair of evaluation scores giving the probability of belonging to the text and a second one of the pair of evaluation scores giving the probability the score showing the possibility of belonging to the text and that of belonging to the mathematical expression;
 - an evaluation unit configured to obtain from the first dictionary the first and second [[the]] evaluation scores showing the probability possibility of belonging to the text and that of belonging to the mathematical expression for each of the words included in the characters recognized by the character recognition unit with reference to the first dictionary; and
 - a mathematical expression detecting unit configured to search for an optimal path connecting words by selecting one of the text and the mathematical expression based on a formative grammar and the first and second evaluation scores showing the possibility of belonging to the text and that of belonging to the mathematical expression for each of the words, the optimal path having the largest sum of evaluation scores

given to a word, thereby detecting characters belonging to the mathematical expression.

2. (Currently amended) The device according to claim 1, wherein said mathematical expression detection unit comprises:

a second dictionary configured to store a connectable [[a]] part of speech and mathematical expression as the formative grammar; and

a search unit configured to search for a path connecting the words and showing the largest evaluation score given to the word as the mathematical expression or the text, out of all possible inter-word connection paths as the optimal path, by selecting either the text or mathematical expression for each word according to the part of speech of the word and the formative grammar read out from said second dictionary.

3. (Currently amended) The device according to claim 1, further comprising:

a memory configured to store a plurality of items of sample information indicating a relation of a normalization size and a center position between each pair of consecutively arranged characters in terms of the types of the characters including a horizontal positional relationship, character/subscript relationship and character/superscript relationship; and

a determination unit configured to calculate the relation of the normalization size and the center position between each pair of consecutively arranged characters included in [[the]] a mathematical expression region and obtain link candidates for the horizontal positional relationship, the character/subscript relationship and the character/superscript relationship based on the calculated relation of the normalization size and the center position, and the sample information coming from the memory.

and corresponding to the calculated relation of the types of the two consecutively arranged characters.

4. (Currently amended) The device according to claim 3, further comprising:
a memory configured to store storing a global evaluation condition [for] determined based on [[the]] a distribution of the heights of the characters contained in said mathematical expression region; [[and]]
a unit [[for]] configured to search, using data from the memory to find [[for]] an optimal path for connecting the characters in each of said mathematical expression regions without contradiction[[,]];
a unit configured to select an inter-character structure candidate having a horizontal positional relationship, a character/subscript relationship or a character/superscript relationship for each pair of consecutively arranged characters based on said global evaluation condition and said link candidates[[,]]; and
a unit configured to recognize the horizontal positional relationship, the character/subscript relationship or the character/superscript relationship of said pair of consecutively arranged characters based on the result of the search operation.

5. (Currently amended) The device according to claim 4, wherein said global evaluation condition comprises at least one of the relationship of the height of a character contained in a subscript region and the height of each of other characters, the positional relationship between a base line and a character contained in the subscript region, and the dispersion of heights among characters located on [[the]] a same horizontal level.

6. (Original) The device according to claim 3, further comprising:

a decomposing unit configured to decompose each mathematical expression detected by said mathematical expression detection unit into components and remove at least left indexes, accent marks, root signs, and dots from each component, and wherein said determination unit obtains link candidates for the components from which the left indexes, accent marks, root signs, or dots is removed.

7. (Canceled).

8. (Canceled).

9. (Currently amended) A mathematical expression recognizing method comprising:

recognizing characters in a document image containing a text and a mathematical expression;

referring to a first dictionary which stores a pair of evaluation scores for each type of word that can be identified by ~~means of~~ normal expression, a first one of the pair of evaluation scores giving the probability of belonging to the text and a second one of the pair of evaluation scores giving the probability the score showing the possibility of belonging to the text and that of belonging to the mathematical expression to obtain the evaluation scores showing the probability possibility of belonging to the text and that of belonging to the mathematical expression for each of the words included in the characters recognized by the character; and

searching for an optimal path connecting words by selecting one of the text and the mathematical expression based on a formative grammar and the first and second evaluation scores ~~showing the possibility of belonging to the text and that of~~

~~belonging to the mathematical expression~~ for each of the words, thereby detecting characters belonging to the mathematical expression.

Claims 10-17 (Canceled).

18. (New) The method according to claim 9, further comprising:
configuring a second dictionary to store a connectable part of speech and mathematical expression as the formative grammar; and
searching for a path connecting the words and showing a largest evaluation score given to the word as the mathematical expression or the text, out of all possible inter-word connection paths as the optimal path, by selecting either the text or mathematical expression for each word according to the part of speech of the word and the formative grammar read out from the second dictionary.

19. (New) The method according to claim 9, further comprising:
storing a plurality of items of sample information indicating a relation of a normalization size and a center position between each pair of consecutively arranged characters in terms of the types of the characters including a horizontal positional relationship, a character/subscript relationship and a character/superscript relationship; and

calculating the relation of the normalization size and the center position between each pair of consecutively arranged characters included in a mathematical expression region and obtain link candidates for the horizontal positional relationship, the character/subscript relationship and the character/superscript relationship based on the calculated relation of the normalization size and the center position, and the sample

information corresponding to the calculated relation of the types of the two consecutively arranged characters.

20. (New) The device according to claim 19, further comprising:
 - storing a global evaluation condition determined based on a distribution of the heights of the characters contained in the mathematical expression region;
 - finding an optimal path for connecting the characters in each of said mathematical expression regions without contradiction;
 - selecting an inter-character structure candidate having a horizontal positional relationship, a character/subscript relationship or a character/superscript relationship for each pair of consecutively arranged characters based on the global evaluation condition and the link candidates; and
 - recognizing the horizontal positional relationship, the character/subscript relationship or the character/superscript relationship of the pair of consecutively arranged characters based on the result of the search operation.

21. (New) The method according to claim 20, wherein said global evaluation condition comprises at least one of the relationship of the height of a character contained in a subscript region and the height of each of other characters, the positional relationship between a base line and a character contained in the subscript region, and the dispersion of heights among characters located on a same horizontal level.

22. (New) The device according to claim 19, further comprising:
 - decomposing each mathematical expression detected into components; and

removing at least left indexes, accent marks, root signs, and dots from each component, wherein link candidates are obtained for the components from which the left indexes, accent marks, root signs, or dots is removed.